Prediction of Myocardial Infarction by using Multi Linear Regression, Neural Network and Logistic Regression Techniques: A Review

RuchiPanwar[1], Amit Kishor[2], Pankaj Pratap Singh[3], Anirudh Kumar tripathi[4]
[1]M.*Tech Scholar,* [234]*Assistant Professor, CSE department, S.I.T.E, Swami Vivekanand Subharti University, Meerut, India*

Abstract- In the field of medical system, myocardial infarction detection has become a laborious task. The massive amount of patient's data keep for detecting the casual of future disease. For the development ofintelligent system would appreciate timely medical care pursued by the proper treatment that is life saving and low cost. For the detection of Myocardial infarction, data mining techniques have been proposed with the changing level of success and accuracy. However, the comparative study play a vital role in the detection of heart disease and it should be done vary practically, absolutely and accurately. The multi linear regression, neural network and logistic regression is used which helps to detect the myocardial infarction and improves the quality of performance of clinical choices.

I. INTRODUCTION

In latest years, accurate detection outcomes using multi linear regression, neural network and logistic regression techniques have been established in many areas of health care research. These method have comparison study and advantages and disadvantages. Myocardial infarction disease is high-spirit disease of the world. Cardiac dataset is also a method which is used to detect or observe cardiac disease and to assess the heart's anatomy. A big problem faced by the health care organization is a foundation of quality services at budget cost.

The clinical outcome is essential to successful decision making for accurate prediction. The aim of this paper is to describe the myocardial infarction detection using multilinear regression neural network and logistic regression techniques. These hidden

Hidden patterns and predictive modeling techniques can be used for health diagnosis in the medicinal data. There are no. of factor which increases the risk of heart disease.

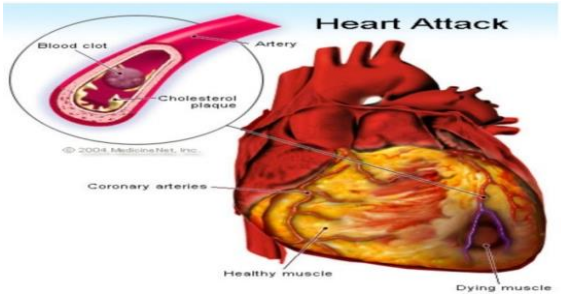

Fig.1:

THERE ARE SOME FACTORS ARE LISTED BELOW:

1) Blood pressure
2) Chest pain.
3) Cholesterol.
4) Fasting blood sugar.
5) Maximum heart rate.
6) Depression.
7)Exercise induced angina.

Over about 85% of death in the today's world are because of myocardial infarction. There are number of difficulties will turn up, when there is a lack of sign, symptoms and physical examination of patient. In recent years, accurate outcome prediction model using multi linear regression, neural network and logistic regression analysis have afore developed in many areas of the health care research. This paper aims to compare the performance and accuracy of the multi linear regression, neural network and logistic regression techniques in prediction of outcomes in myocardial infarction. Multi linear regression, neural network and logistic regression techniques have been applied by the researchers in various medical areas and also have been used for estimating risk in the myocardial infarction disease and logistic regression are also used to evaluate disease risk in coronary heart disease. In this research, the comparative analysis of the multi linear regression, neural network and logistic regression to be useful tools in medical diagnosis to improve the accuracy in the efficient manner.

II. LITERATURE REVIEW

1) Subhadra et al. ".Neural networks based intelligent system for prediction heart disease". Back propagation algorithm was applied to train the data and compare the parameters iteratively and the proposed system of heart disease prediction with appropriate diagnosis have been framed up using multilayer perception neural network.

2) Reddy et al. proposed work divided into 2 parties performance model and prediction model. Performance model is designed to evaluate the overall performance of application. Prediction model is used to predict the pressure cholesterol etc. The accuracy of system is proved in java.

3)Hazra et al. summarized some of the current research on predicting heart disease using data mining technique , analyze the

various combination of mining algorithm used and conclude which technique are effective and efficient.

4)Sayadet al. presented diagnosis system for heart disease using neural network approach data mining and Artificial neural network techniques a multilayer perception neural network along back propagation algorithm is used to develop the system because MLPNN model proves the better result and helps the domain experts

MULTI LINEAR REGRESSION

Multi linear regression is a statistical model that can be used to explain the relationship between one dependent variable and two or more independent variable and also used to describe the data. Analyzing the co relation directionality of the data, fit in the line and evaluating the validity and usefulness of the model are different stages of multiple linear regression model. The regression line represent the estimated disease chance for a given combination of input factor B. Scatter plot is defined by the linear equation

$y=\beta0 + \beta1\ x1+ \beta2x2+.....+ \beta ixn$ for $i = 1\ldots n$

To minimize the residual, the method of least squares used. The multi linear regression model is fitted efficiently R square is calculated as

Rsquare = SSR/SST =sum of squares/total sum of squares

NEURAL NETWORK:

A neural network is a sequence of algorithm that whirl to recognize underlying relationship in a set of data through a process that imitate the way the human brain operate. Neural network are the computer models which pretend neural process by summing inhibitory and excitatory inputs to produce the output criteria and the network generate best possible result without needing to redesign or reconstruct output.

Neural Network a simple computational task generally a basic yes/no decision or on/off decision and they consist of highly interconnected nodes and each nodes in the input layer is called input nodes and represents an input variable i.e. used as predictor of the outcome [13]

ARCHITECTURE OF NEURAL NETWORK:

1) Input Layer: The input layer contain three artificial neurons which are to receive input from the outside world. This is where the actual learning on the neural network happens or all the input are fed in the model by this layer.

2) Output Layer: The output layer contain units that respond to the information i.e. fed into the system and also whether it learned and task or not.

3) Hidden Layer: The hidden layer are described hidden in between input layer and output layer. The only job of a hidden layer is to transform the input layer into something meaningful that the output layer can use in same way.

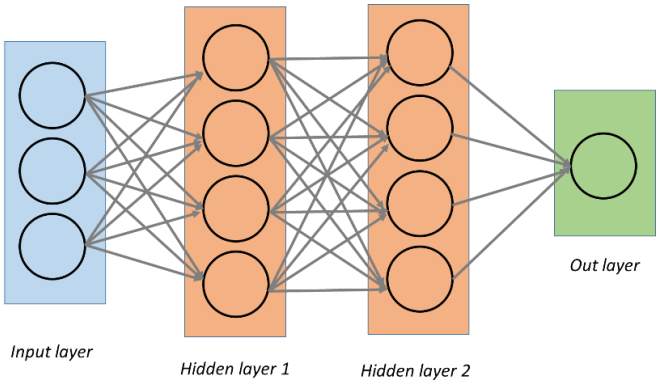

Fig.2:

RISK FACTOR FOR HEART DISEASE:

1) Blood pressure: High BP leads to heart attack.

2)Stress: stress can lead to heart attack and strokes

3)Cholesterol: to reduce the risk of heart disease a saturated transfat will low cholesterol level.

4)Smoking: Smokers risk a heart attack thrice as much as non smokers

5)Diabetes: If human couldn't control then diabetes can lead to heart attack and death.

LOGISTIC REGRESSION TECHNIQUE:

Logistic regression describe the relationship between a binary (dependent) variable such as presence or absence of disease and predictor (independent) variables such as patient demographics or imaging findings. For example, the presence or absence of heart attack within a specified time period might be predicted from knowledge of the patient's age, cholesterol level, family history of heart attack, and any prior attack procedures. The outcome variables can be both continuous and categoric. If $X_1, X_2,\ldots, X_n$ denote n predictor variables (e.g, calcification types, cholesterol level, patient age, and so on), Ydenotes the presence ($Y = 1$) or absence ($Y = 0$) of disease, and p denotes the probability of disease presence (ie, the probability that $Y = 1$), the following equation describes the relationship between the predictor variables and p:

$\text{Log}(p1-p)=\beta0+\beta1X1+\ldots+\beta nXn,$

where β_0 is a constant and $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients of the predictor variables $X_1, X_2, \ldots, X_n$. The regression coefficients are estimated from the available data. The probability of disease presence p can be estimated with this equation.[12]

Each regression coefficient describes the size of the contribution of the corresponding predictor variable to the outcome. The effect of the predictor variables on the outcome variable is commonly measured by using the odds ratio of the predictor variable, which represents thefactor by which the odds of anoutcome change for a one-unit change in the predictor variable. The odds ratio is estimated by taking the exponential of the coefficient (e .g, $\exp[\beta_1]$). For example, if β_1 is the coefficient

of variable X_{FH} ("family history of heart attack"), and p represents the probability of heart attack, $\exp(\beta_1)$ is the odds ratio corresponding to the family history of heart attack. The odds ratio in this case represents the factor by which the odds of having heart attack increase if the patient has a family history of heart attack and all other predictor variables remain unchanged. In other words, if the odds ratio corresponding to the family history of heart attack is 2, then heart attack occurs twice as often in men with a family history of heart attack in comparison with men in the study population with no such family history[12].

Logistic regression models generally include only the variables that are considered "important" in predicting an outcome. With use of P values, the importance of variables is defined in terms of the statistical significance of the coefficients for the variables. The significance criterion $P \leq .05$ is commonly used when testing for the statistical significance of variables; however, such criteria can vary depending on the amount of available data. For example, if the number of observations is very large, predictors with small effects on the outcome can also become significant. To avoid exaggerating the significance of these predictors, a more stringent criterion (e.g, $P \leq .001$) can be used.

Significant variables can be selected with various methods. In forward selection, variables are sequentially added to an "empty" model (ie, a model with no predictor variables) if they are found to be statistically significant in predicting an outcome. In contrast, backward selection starts with all of the variables in the model, and the variables are removed one by one as they are found to be insignificant in predicting the outcome. The stepwise logistic regression method is a combination of these two methods and is used to determine which variables to add to or drop from the model in a sequential fashion on the basis of statistical criteria. Although different techniques can yield different regression models, they generally work similarly. Sometimes, clinically important variables may be found to be statistically insignificant with the selection methods because their influence may be attenuated by the presence of other strong predictors. In such cases, these clinically important variables can still be included in the model irrespective of their level of statistical significance. [12].

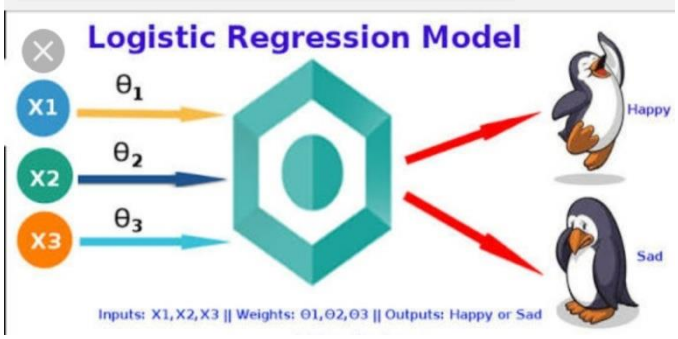

Table: Comparison of all three predicting techniques

Neural Network	Multi Linear Regression	Logistic Regression
Less statistical knowledge are required for model building in neural network	Statistical knowledge are required for model building in neural network	More statistical knowledge are required for model building in logistic regression
There is automatically ability to detect the complex relationship between input and output	There is an ability to detect the complex relationship between input and output	There is no ability to detect the complex between input and output
Implicit ability to detect interaction	explicit ability to detect interaction	explicit ability to detect interaction.
Neural network uses softmax function	Multi linear regression uses anova function	Logistic regression uses sigmoid function
More computational time	More computational time	Less computational time
Confidence interval are difficult to calculate	Confidence interval are normal to calculate	Confidence intervals are easy to calculate
More robust	Multi linear regression are robust in nature.	Less robust
Good discrimination ability	Less discrimination ability	Good discrimination ability
Prone to overfiting	Prone to overfiting	Less overfiting

III. CONCLUSION

Heart disease have increased a lot in recent times. Due to the increasing amount of hearts patient , It is difficult to maintain the massive amount of data in hard copy. Indeed a better way is to store the massive amount of patient's data stored for predicting the casual of future disease. The main purpose of this research is to carry the information about patient's classification and regression use it for detecting the myocardial infarction. The comparative study plays a important role in this detection of heart diseases. It should be done very effective , precisely and accurately.

IV. REFRENCE

[1]. K.Subhadra , Vikas B ".Neural networks based Intelligent system for prediction heart disease". ISSN-2278-3075.(2019).

[2]. Abdul Ghaffar Khan "Prediction of heart disease using artificial neural network. (2018).

[3]. Sofia Tomov "For deep neural network for detecting heart disease"(2018).

[4]. K sandhya rani , M Sai Manoj , G. Sugumani "A heart disease prediction using heart disease prediction using logistic regression (2018).

[5]. Kartik Chopra , C. Srimathi "Logistic regression and CNN performance analysis based on the size of the dataset. ISSN -2321-9939" (2018).

[6]. AnimeshHazra , Subrata Kumar Mandal , Amit Gupta , Arkomita Mukherjee and Asmita Mukherjee to summarize some of the current research on predicting heart disease using data mining technique. ISSN-0973-6107 (2017).

[7]. K.Polaraju, D.Durga Prasad "Prediction of heart disease using milti linear regression model". ISSN-2321-993.(2017).

[8]. Aarti Londhe , Prof. AmolJoglekar "Analyzing the risk of heart disease using ANN".ISSN-2278-1323.(2016).

[9]. Ebenezer Olaniyi , Adnan Khashman, OyebadeK.Oyedotun "Heart disease diagnosis using neural network arbitration (2015)

[10]. A.T.Sayad , P.P.Halkarnikar , presented diagnosis system for heart disease using neural network approach data mining and Artificial neural network techniques. ISSN-2321-9009(2014)

[11]. S.Florence , N.G BhuvaneshwarAmma . G. Anna poorani , K malathi " Predicting the risk of heart attack using neural network and decision tree. (2014).

[12]. http://pubs.rsna.org/.doi/full/10.1148/rg.301095057

[13]. http://mindmajix.com/artificial-neural-network-and-- How-it-works

[14]. M.Akhil Jabbar , B.L.Deekshatulu and Preeti Chandra "Classification of heart disease using artificial neural network and feature subset selection" ISSN-0975-4172(2013).

[15]. Chaitrali Dangare (A data mining approach for prediction of heart disease using neural network)(2012).

Authors' Profile

Ruchipanwar is pursuing M. Tech. from Subharti Institute of Engineering and Technology, Swami VivekanandSubharti University, Meerut, India. She received her B. Tech Degree in computer science and Engineering from Uttar Pradesh Technical university, Lucknow, India. Her area of interest is regression.

Er. Amit Kishor is working as Assistants Professor in the department of Computer Science Engineering and I.T., Subharti Institute of Engineering and Technology, Swami VivekanandSubharti University, Meerut, India. Currently he is pursuing Ph. D. in Computer Engineering from Department of Computer Science and I.T., Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad

Er. Pankaj Pratap Singh received his B. Tech Degree in Computer Science Engineering from Uttar Pradesh Technical University, Lucknow, India, in 2007 and M. Tech degree in Medical Image and Image Processing from Indian Institute of Technology Kharagpur, Kharagpur, India, in 2010. He is currently working as Assistant Professor in the Department of Computer Science Engineering and Information technology, Subharti Institute of Engineering and Technology, Swami Vivekanand University, Meerut, India. His research interests include IOT Neural

Er. AnirudhkumarTripathi is working as Assistant Professor in the department of Computer Science Engineering and I.T., Subharti Institute of Engineering and Technology, Swami VivekanandSubharti University, Meerut, India.